2005

Scalable multi-resolution color image segmentation

Fardin Akhlaghian Tab
University of Wollongong

Golshah Naghdy
University of Wollongong, golshah@uow.edu.au

A. Mertins
University of Wollongong, mertins@uow.edu.au

Scalable multiresolution Color Image Segmentation

Farin Akhlaghian Tab[a], Golshah Naghdy[a] and Alfred Mertins[b]

[a]School of Electrical, Computer and Telecommunications Engineering
University of Wollongong, Wollongong, Australia;
[b]Signal Processing Group, Institute of Physics
University of Oldenburg, Oldenburg, Germany

ABSTRACT

This paper presents a novel multiresolution image segmentation method based on the discrete wavelet transform and Markov Random Field (MRF) modelling. A major contribution of this work is to add spatial scalability to the segmentation algorithm producing the same segmentation pattern at different resolutions. This property makes it applicable for scalable object-based wavelet coding. To optimize segmentation at all resolutions of the wavelet pyramid, with scalability constraint, a multiresolution analysis is incorporated into the objective function of the MRF segmentation algorithm. Examining the corresponding pixels at different resolutions simultaneously enables the algorithm to directly segment the images in the YUV or similar color spaces where luminance is in full resolution and chrominance components are at half resolution. In addition to spatial scalability, the proposed algorithm outperforms the standard single and multiresolution segmentation algorithms, in both objective and subjective tests, yielding an effective segmentation that particularly supports scalable object-based wavelet coding.

Keywords: Segmentation, Scalability, Markov Random Field variable

1. INTRODUCTION

Image segmentation is the process of dividing an image into homogenous regions, which is an essential step toward higher level image processing such as image analysis, pattern recognition and computer vision. In particular, effective segmentation is crucial for the emerging object-based image/video standards such as object-based coding standards defined by MPEG-4[1] and content-based shape descriptor used in MPEG-7.[2]

In scalable object-based coding, a single codestream can be sent to different users with different processing capabilities and network bandwidths by selectively transmitting and decoding the related parts of the codestream.[3] Some of the desirable scalable functionalities are signal-to-noise ratio (SNR) scalability, spatial scalability and temporal scalability.[3] A scalable bitstream includes embedded parts that offer increasingly better SNR, greater spatial resolution or higher frame rates. Therefore considering the spatial scalability, it is necessary to extract and present objects' shape at different resolutions for the scalable object-based encoder/decoder systems. For an effective scalable object based coding algorithm, it is desirable that the shapes of the extracted objects at different resolutions be similar or equivalently, the pattern of segmented regions should be similar at different resolutions. We call the segmentation algorithm with the similar patterns at different resolutions scalable segmentation.

Multiresolution image segmentation algorithms analyzes the image data at different resolutions which results in some advantages compared to single resolution segmentation such as : less computational complexity, improved convergence rate, less oversegmentation, less sensitivity to noise, ability to capture the image structures at different resolutions and less dependence on initial segmentation. These algorithms consider the inter scale image data correlation in the segmentation procedure. In the most straight forward case, these algorithms consider inter scale correlation by projection of the lower resolution segmentation result to the next higher resolution as an initial segmentation estimation. The segmentation is further refined at the current higher resolution by a single resolution segmentation. This procedure continues progressively until the highest resolution is segmented,[4,5,6,7]

In a second group of segmentation algorithms, the inter scale correlations are considered in the statistical models and decision at each pixel/block is based on the information of the different resolutions,[8,9,10] However,

Visual Communications and Image Processing 2005, edited by Shipeng Li,
Fernando Pereira, Heung-Yeung Shum, Andrew G. Tescher, Proc. of SPIE Vol. 5960
(SPIE, Bellingham, WA, 2005) · 0277-786X/05/$15 · doi: 10.1117/12.633217

often only the causal inter scale correlation with the next lower resolution,[8] [9] or the next higher resolution is considered.[10] Considering the other resolutions results in a very complex model and increases the computational complexity.

None of the known works in the literature have considered interscale correlation between all pyramid resolutions. In addition their extension to scalable segmentation, producing the same segmentation patterns at different resolution, is nearly impossible or results in an algorithm with large computational complexity. In order to produce similar objects/regions at different resolutions, we present a novel MRF based multiresolution grey/color image segmentation algorithm which extends the statistical model in order to consider the correlation between all the resolutions without overly increasing the computational complexity. It produces the same segmentation patterns at different resolutions and is applicable to object-based wavelet coding algorithms. Bearing in mind the downsampling of pixels to lower resolution in wavelet decomposition, corresponding pixels at different resolutions are considered as vector of pixels which are supposed to have the same segmentation label. A vector or multiresolution/multidimensional analysis is incorporated in the objective function of the MRF-based segmentation algorithm which aligns the segmentation algorithm with the object-based wavelet coding spatial scalability. This vector analysis keeps wavelet spatial scalability as a constraint which fits multiresolution MRF segmentation to wavelet-based scalable object coding.

The multiscale analysis uses different-resolution information concurrently, which produces better results than regular single and multiresolution Bayesian segmentation algorithms. It combines and processes both coarse global information of low resolution with fine local information obtained from higher resolutions of the wavelet decomposition pyramid. Therefore it combines good features of both single and multiresolution segmentations. While it is noise resistant, it detects objects/regions better than regular multiresolution segmentation and also results in a lower number of regions than single-level segmentation. To optimize the objective function of the segmentation algorithm, the Iterated Condition Mode (ICM) algorithm according to,[4] matched to the scalable multiscale analysis, is used.

Color images have more information than grey-level images which results in more reliable separation of foreground regions from background in color image segmentation algorithms. It has been recognized that selection of an appropriate color space produces more perceptually effective segmentation results,[11] .[12] In particular, segmentation in YUV or LUV spaces produces more favorable results than the RGB space,[11] ,[12] .[13] Many of the images and image sequences in the databases are in YUV format where Y is in full resolution while U and V components are in half resolution. The fact that the Y, U, and V channels are presented at different resolutions is not considered in any of the the existing regular single or multiresolution color image segmentation algorithms. However, this fact calls for a fitted multiresolution algorithm to perform the segmentation task effectively. The proposed algorithm has enough flexibility to directly segment color images. In the vector analysis, only available components of color data at different resolutions are used to classify the vector to one of the segmentation labels. For this reason, the proposed algorithm can segment grey-level images as well.

In order to produce similar objects/regions at different resolutions, an alternative method is single-resolution segmentation followed by downsampling. In single-resolution region-based image/video segmentation algorithms, features such as intensity/color, texture, motion, etc. are considered at the highest resolution. In this method, good features of multiresolution segmentation such as reduced noise sensitivity are lost, and producing optimized and visually pleasing objects/regions at different resolutions, as a criterion, is not considered. Furthermore, downsampling distorts shapes and cannot preserve their topology at lower resolutions for all possible shapes.[14] In other words, achieving visually pleasing objects/regions at higher resolution does not necessarily ensure similar quality at lower resolutions. For example in Figure 1, downsampling of two digital circles is compared. It can be seen that better approximation of a digital circle at high resolution can result in worse downsampled shape.

This paper is organized as follows. Section 2 refers to the scalability in object-based wavelet coding. In Section 3, the proposed scalable multiresolution segmentation algorithm, which includes a statistical image modeling and optimization processes, is explained. Some experimental results and discussion are presented in Section 4, and finally, conclusions are drawn in Section 5.

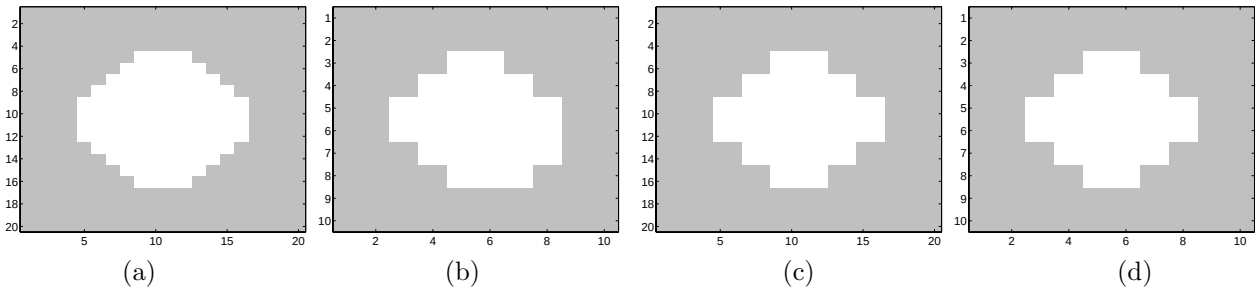


Figure 1. Circles in different resolutions. (a) closer approximation of a digital circle at High resolution; (b) downsampling to low resolution; (c) worse approximation of a digital circle at high resolution; (d) downsampling of (c) to low resolution.

2. OBJECT-BASED WAVELET CODING SCALABILITY

Scalability means the capability of decoding a compressed sequence at different data rates. It is useful for image/video communication over heterogenous networks which require a high degree of flexibility from the coding system. Some of the desirable scalable functionalities are signal-to-noise ratio (SNR) scalability, spatial scalability and temporal scalability.[3] In particular spatial scalability means that, depending on the end user's capabilities (bandwidth, display resolution etc.), a resolution is selected and all the shape and texture information is sent and decoded at the appropriate resolution. Scalable image/video coding has also different applications such as web browsing, image/video database systems, video telephony, etc.

In wavelet-based spatial scalability applications, due to the self similarity feature of the wavelet transform, the shape in lower scale is the shape in the lowpass (LL) subband. The exact relationship between the full-resolution shape and its low-resolution versions depends on the kind of wavelet transform used for the decomposition. In this paper we use an odd length filter (e.g. 9/7), where all shape points with even indices* are downsampled for the lowpass band.[15] Figure 2 further illustrates the wavelet decomposition of arbitrarily shaped objects when using an odd-length filter. The final four-band decomposition is depicted in Figure 2(c). As a result, every shape pixel with even indices has a corresponding pixel on the lower resolution and every shape pixel on the lower level has a corresponding pixel on the next higher level. By considering the self similarity of the wavelet transform, it is straightforward to suppose that the pixels of a shape with even indices have the same segmentation classifications as the corresponding pixels on the lower level.

The wavelet self similarity extends to all low pass subband shapes of different levels. Therefore the discussed relationship between corresponding pixels is extended to shapes at different scales. Corresponding pixels at different resolutions have the same segmentation class.

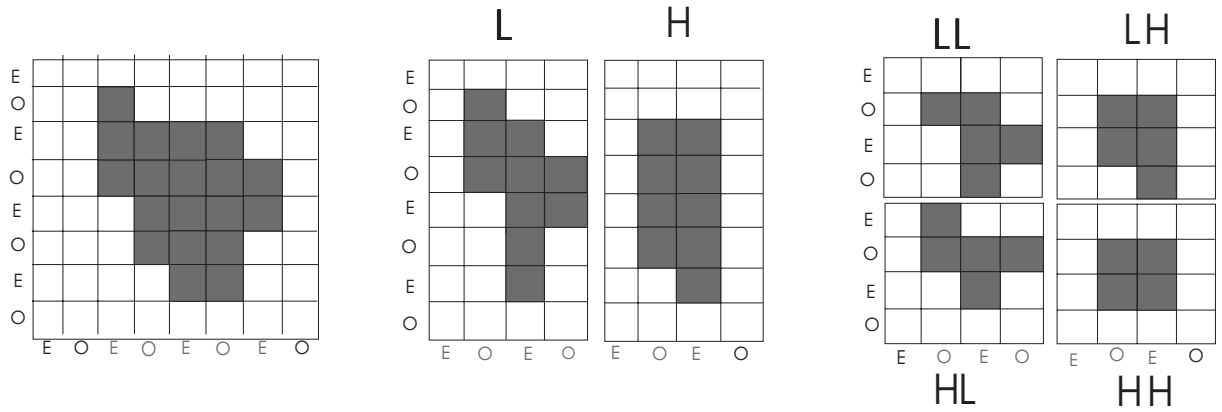
3. SPATIAL SEGMENTATION ALGORITHM

Markov Random Field statistical modeling is used in many image processing applications. In order to solve an image processing problem by the MRF technique, a statistical image model has to be fitted to the application which captures the intrinsic character of the image in a few parameters. Image/Video processing problems, including all uncertainties and constraints, can therefore be converted to a mathematical parameter optimization problem.[16]

3.1. Statistical color image model

The main challenge in multiresolution image segmentation for scalable object-based wavelet coding is to keep the same relation between extracted objects/regions at different resolutions as it exists between the decomposed objects/regions at different resolutions in a shape adaptive wavelet transform. The other constraint is border smoothness particularly in lower resolutions. Different smoothness coefficients defined at different resolutions give some degree of freedom to put more emphasis on the low-resolution smoothness. To meet these challenges,

*Suppose indices start from zero or an even number.



(a) (b) (c)

Figure 2. Decomposition of a non rectangular object with odd-length filters; (a) the object, shown in dark gray; (b) the decomposed object after horizontal filtering; (c) decomposed object after vertical filtering. The letters "E" and "O" indicate the position(even or odd) of a pixel in the horizontal and vertical dimensions.

Markov random field modeling is selected as it includes low level processing at pixel level and has enough flexibility in defining objective functions matched with the problem at hand.[16] We first explain the statistical model of single resolution grey/color image segmentation and then extend it to the scalable multiresolution segmentation mode. In a regular single-level MRF-based image segmentation the problem is formulated using a criterion such as the maximum a posteriori (MAP) probability. The desired segmentation is denoted by X, and Y is the observed color image with three channels shown by a three dimensional vector $Y = [Y_1, Y_2, Y_3]$. Then according to the Bayes rule, the a posteriori probability density of the segmentation variables can be written as

$$P(X|Y) \propto P(Y|X)P(X) \tag{1}$$

where $P(X|Y)$ represents the conditional probability of the segmentation label, given the observation Y. By assuming the conditional independence of the channels given the segmentation field,[11] we have $P(Y|X) = P(Y_1|X)P(Y_2|X)P(Y_3|X)$, and the conditional probability in (1) becomes

$$P(X|Y) \propto P(Y_1|X)P(Y_2|X)P(Y_3|X)P(X). \tag{2}$$

The label field X is normally modeled by a MRF stochastic variable. Spatial continuity is easily incorporated into the segmentation, because it is inherent to MRFs.[17] Using a four or eight pixel neighborhood system considering only pairwise cliques, $P(X)$ is then a Gibbs distribution[4] and is defined by its energy function $U(X)$ such that

$$P(X) = \frac{1}{Z} \exp\left(-\frac{1}{T}U(X)\right), \quad U(X) = \sum_{c \in C} V_c(X) \tag{3}$$

where C is the set of all cliques, and V_c is the clique potential function. A clique is a set of neighboring pixels. A clique function depends only on the pixels that belong to the clique. In single-resolution segmentation, usually one or two pixel cliques are used as shown in Figure 3(a), and for one pixel cliques we assume that the one pixel clique potentials are zero, which means that all region types are equally likely.[4] Spatial connectivity of the segmentation is imposed by assigning the following clique function:

$$V_c(s,r) = \begin{cases} -\beta & \text{if} \quad X(s) = X(r) \\ +\beta & \text{if} \quad X(s) \neq X(r) \end{cases}, \qquad (s,r) \in \mathrm{C} \tag{4}$$

Herein β is a positive number and s and r are a pair of neighboring pixels. Note that a low potential or energy corresponds to a higher probability for pixel pairs with identical labels and lower probability for pairs with different labels, which automatically encourages spatially connected regions.

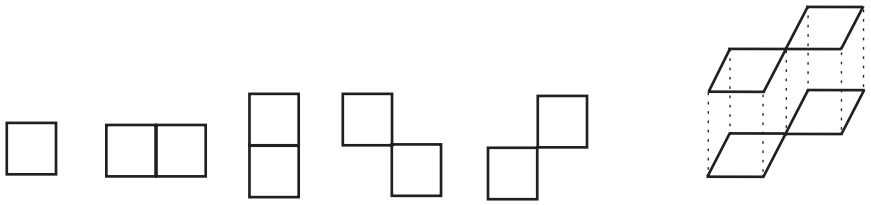


Figure 3. (a) Normal one and two pixels cliques sets. (b) A clique of two vectors with the vectors' dimension equal to two.

The conditional probability density $P(Y_i|X),\ i = 1, 2, 3$ is modeled as a white Gaussian process, with mean $\mu^i_{X(s)}(s)$ and variance σ_i^2 for channel i. Each region is characterized by a mean vector $[\mu^1_{X(s)}(s),\ \mu^2_{X(s)}(s),\ \mu^3_{X(s)}(s)]$ which is a slowly varying function of s. Therefore $P(Y|X)$ can be described by the following equation:

$$P(Y|X) \propto \exp\ \left\{\ -\sum_s \left(\sum_{i=1}^{3} \frac{1}{2\sigma_i^2}(Y_i(s) - \mu^i_{X(s)}(s))^2 \right) \right\} \tag{5}$$

Considering equations (1), (3) and (5), the conditional probability density of the segmentation variable becomes:

$$P(X|Y) \propto \exp\ \left\{\ -\sum_s \left(\sum_{i=1}^{3} \frac{1}{2\sigma_i^2}(Y_i(s) - \mu^i_{X(s)}(s))^2 + \frac{1}{T} \sum_{r \in \partial_s} V_c(s, r) \right) \right\} \tag{6}$$

It is easy to see that the parameters $\sigma_i,\ i = 1, 2, 3$, T and β are interdependent. Therefore, to simplify the expression, the parameters $2\sigma_i^2, i = 1, 2, 3$ and T are set to one, and the segmentation result is controlled by the value of β in the V_c function. The probability density function has two components. One forces the region intensity to be close to the data, and the other imposes spatial continuity. Considering the MAP criterion, we maximize the probability $P(X|Y)$, which is equivalent to minimizing the negative value of argument of the exponential function in equation (6). This results in the following cost or objective function which has to be minimized with respect to $X(s)$:

$$E(X) = \sum_s \left(\sum_{i=1}^{3} (Y_i(s) - \mu^i_{X(s)}(s))^2 + \frac{1}{T} \sum_{r \in \partial_s} V_c(s, r) \right) \tag{7}$$

In grey-scale images only the intensity channel exists, and the terms Y_1 and Y_2 in equation (7) have to be deleted.

To obtain the final segmentation, this objective function is minimized by one of the several MRF objective minimization methods.[16] To tailor this objective function to scalable multiresolution color image segmentation, initially, the wavelet transform is applied to the original image and a pyramid of decomposed images at various resolutions is created. Let $Y = [Y_1, Y_2, Y_3]$ where $Y_i, i = 1, 2, 3$ is the intensity of channel i of the pyramid's pixels. The segmentation of the image into regions at different resolutions will be denoted by X.

As mentioned earlier, considering scalability, a pixel and its corresponding pixels at all other pyramid levels have the same segmentation label. Therefore they change together during the segmentation process. To change the segmentation label of a pixel, the pixel and all its corresponding pixels at all other levels have to be analyzed together. As a result, an analysis of a set of pixels in a multidimensional space instead of a single resolution analysis needs to be used. The term "vector" is used to refer to multidimensional space. A vector includes corresponding pixels at different resolutions of the pyramid. A symbol $\{s\}$ shows a vector which includes pixel s. The dimension of a vector is equal to the number of it's pixels which are located at different resolutions. If the corresponding pixels are determined according to the wavelet transform downsampling, the vector dimension depends on the index of pixels, and it can be 1, 2 or more. Using these primary definitions, the clique concept

is extended to vector space. The extended cliques act on two vectors instead of two pixels. Figure 3(a) shows regular one and two pixels clique sets. In Figure 3(b), the extension of one of these cliques to the array mode in two dimensional space can be seen.

The extension of clique functions is achieved through the following steps: equation (4) is used for cliques of length two at a resolution where pixels s and r are two neighboring pixels at the same resolution level. Equation (8) below is defined for multiple levels, where $\{s\}$ and $\{r\}$ are vectors corresponding to two neighboring pixels s and r. The neighboring pixels of the two vector $\{s\}$ and $\{r\}$ at level k are denoted as s_k and r_k. The lowest resolution which include a pixel of vector $\{s\}$ is denoted as M and N is the dimension of vectors $\{s\}$ and $\{r\}$. A positive value is assigned to the parameter β, so that adjacent pixels, of two neighboring vectors, are more likely to belong to a same region than to different regions. Increasing the value of β decreases the sensitivity to intensity changes.[4]

$$V_c(\{s\},\{r\}) = (\frac{1}{N}) \sum_{k=M}^{M+N-1} (-1)^{L_k}.\beta, \quad L_k = \begin{cases} 1 & \text{if } X(s_k) = X(r_k) \\ 0 & \text{if } X(s_k) \neq X(r_k) \end{cases} \quad s_k \in \{s\},\ r_k \in \{r\},\ r_k \in \partial s_k \tag{8}$$

It is notable that the equation (8) extends the clique definition to multiresolution mode. In vector $\{S\}$ corresponding pixels at different resolutions are determined

according to the downsampling relationship of arbitrary shape wavelet transform. In other words the pixels of the lower resolution occur at downsampled positions of the higher-resolution pixels in the pyramid. However, pixels of $\{r\}$ are neighbors of $\{s\}$ at different resolutions. As a result of the clique extension to multiresolution space, segmentation processing will continue in the vector space, therefore, intensity average and segmentation label functions are also extended to vector space. The intensity of pixels at different channels in set $\{s\}$ form a vector $Y(\{s\}) = [Y_1(\{S\}), Y_2(\{S\}), Y_3(\{S\})]$, and similarly, $\mu(\{s\}) = [\mu_1(\{S\}), \mu_2(\{S\}), \mu_3(\{S\})]$ is the mean vector. Therefore the objective function is extended to vector space as follows

$$E(X) = \sum_{\{S\}} \left\{ \sum_{i=1}^{3} ||Y_i(\{s\}) - \mu^i_{X(\{s\})}(\{s\})||^2 + \sum_{\{r\}\in\partial\{s\}} V_c(\{s\},\{r\}) \right\} \tag{9}$$

The outer summation is over vectors, while the first inner summation is related to the distances of the pixel's intensities from the estimated average for each channel of color images. The second inner summation is over all neighborhood vectors of vector $\{s\}$. The two vectors $\{s\}$ and $\{r\}$ are neighbors if pixels of $\{s\}$ and $\{r\}$ located at the same resolution are also neighbors. The approach used in this section, expressed by equations (7) to (9), is a generalization of regular single-resolution segmentation to scalable multiresolution segmentation.

3.2. MAP estimation

The Iterated Condition Mode (ICM) optimization method[18] is used to minimize the objective function in equation (9). The segmentation is initialized with the k-means clustering algorithm for each channel separately. Then neighboring pixels with equal labels at all three channels form a region. The segmentation estimation is improved using ICM optimization.[18] ICM, as used in the single-level segmentation algorithm for grey-level images by Pappas,[4] and extended to color images by M.Chang et al[11] is modified to adapt to the scalable multiresolution segmentation algorithm. Thereofore the objective function term corresponding to the current vector is optimized given the segmentation at all other vectors of the pyramid. The resulting objective function terms related to the current vector are:

$$E(X\{s\}) = \sum_{i=1}^{3} ||Y_i(\{s\}) - \mu^i_{X(\{s\})}(\{s\})||^2 + \sum_{\{r\}\in\partial\{s\}} V_c(\{s\},\{r\}) \tag{10}$$

For grey-level images there is only the intensity channel and the objective function is simplified to

$$E(X\{s\}) = (Y(\{s\}) - \mu_{X(\{s\})}(\{s\})^2 + \sum_{\{r\}\in\partial\{s\}} V_c(\{s\},\{r\}) \tag{11}$$

During the optimization process for each pixels s of a vector $\{s\}$, the terms $\mu^i(s), i = 1, 2, 3$ are estimated by averaging the channel intensities of all pixels that belong to the region i and are inside a window with width w centered at pixels s. The window size w is doubled when we move to the next higher resolution. The average of any pixel s and its correspondences at all other levels in $\{s\}$ are used to classify the pixels of $\{s\}$ to a label which minimizes equation (10). To reduce computational complexity, it is enough to consider only labels of $\{s\}$ and its neighboring vectors to select the best label by the energy minimization through equation (10). Therefore for the pixels inside a region there is no computation and the regions' borders are gradually refined. Furthermore, this border processing prevents isolated noise pixels from becoming new clusters, resulting in fewer wrongly detected boundaries.[19]

Let us consider the overall optimization algorithm now. As mentioned above, the initial segmentation of the pyramid is obtained by the k-means clustering algorithm.[20] The pyramid's pixels, are processed progressively from low to high resolutions. At each resolution, pixels are visited in a raster scan order. The intensity average $\mu^i(s),\ i = 1, 2, 3$ at each pixel s and its corresponding pixels at the other resolutions for all possible classes are estimated with a pre-determined window size w used for estimation. We then update the estimate of $X\{s\}$ using the ICM approach with a multi-level analysis using equation (11). By updating the segmentation labels of pixels at the current resolution, the corresponding pixels at the other levels are also updated. After convergence at the current resolution, the algorithm moves to the next higher resolution and updates the estimates of μ and X and so on, until all resolutions are processed. The stopping criterion at each resolution is the number of X update which should be below a pre-defined threshold. To reduce the number of iterations, other convergence criteria can also be used. The whole procedure is repeated with a smaller window size. The algorithm stops when the minimum window size for the lowest level is reached. We have considered the minimum window size being 8 for the lowest level.

4. EXPERIMENTAL RESULTS AND DISCUSSION

In this section, experimental results obtained using the algorithm introduced in Section 3 are presented. The results are compared with regular single-level and multiresolution segmentation algorithms,[4] .[11] In the first step, the image is decomposed into three resolutions, using the (9/7) wavelet filter. Then in each level of the decomposition, the image is segmented while scalability between regions in different resolutions, as required for the arbitrary shape wavelet transform, is achieved with the proposed algorithm.

As a first example, the proposed algorithm is tested using frame 5 of the SIF sequence Table Tennis. For grey-level images only the intensity of the grey channel is considered in equation (11). Figure 4 represent the results achieved by the proposed multiresolution scalable, regular single and multiresolution segmentation algorithms. The result of the proposed scalable multiresolution segmentation algorithm is presented only at the finest resolution, because the lower resolutions results have the same patterns and figures.

In regular multiresolution segmentation algorithms, brief, coarse and filtered versions of the image are processed at the lower resolutions, therefore, some small-size or low-contrast regions are not detected. This drawback is called undersegmentation. In contrast, in the proposed algorithm the effects of high resolutions on low-resolution results in the detection of details, which is not possible using regular multiresolution segmentation. In other words, the sensitivity to grey-level changes is increased, resulting in a better detection of small or low-contrast objects especially in low resolutions. Table 1 shows the number of detected regions of the Table Tennis image in three spatial resolutions for different segmentation algorithms. The proposed scalable segmentation detects more relevant regions than the regular multiresolution algorithm. For example, consider the segmentation of the textured wall and detection of the ball in the Table Tennis image as presented in Figure 4 by the proposed multiresolution scalable, the regular single-resolution, and the multiresolution segmentation algorithm. The single-level segmentation detects the ball, but it also detects a number of spurious regions due to the textured background as the number of regions in Table 1 shows. This drawback is called oversegmentation. The regular multiresolution algorithm misses the ball at different resolutions altogether. The proposed algorithm, however, detects the ball as well as avoiding unsightly segmentation of the textured background.

It is significant to note that while our algorithm has improved sensitivity to grey-level variation it still maintains noise tolerance. To test the scalable segmentation algorithm on noisy images, first a uniform noise

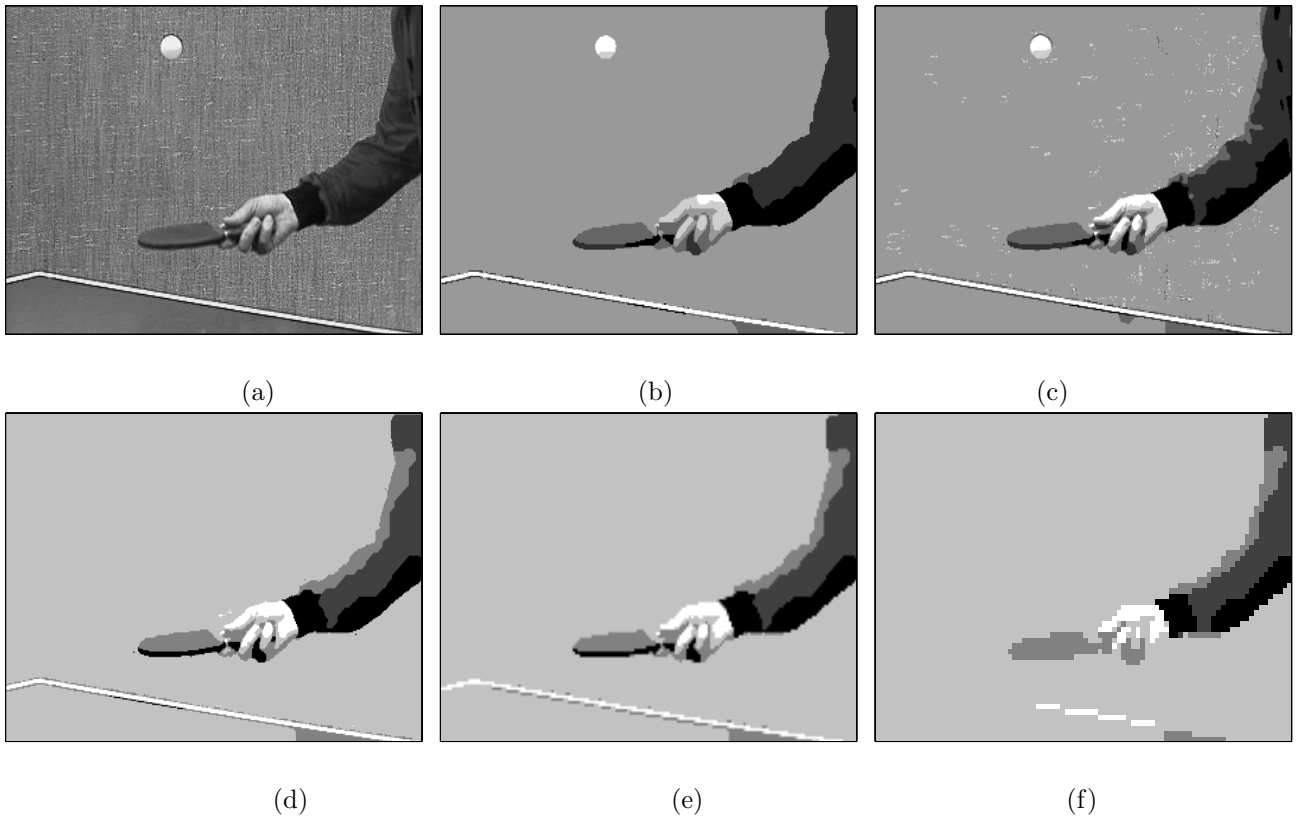


Figure 4. Table tennis image segmentation with $k = 6$ clusters and $\beta = 100$; (a) the main image; (b) segmentation by the proposed scalable algorithm at 240×352; (c) regular single level segmentation; (d) regular multiresolution segmentation 240×352; (e) 120×176 segmentation; (f) 60×88 segmentation;

Table 1. Number of regions in Table Tennis segmentation.

Seg. algorithm	60×120	120×176	240×352
Multi Resolution	19	55	164
Scalable	42	83	184
Single level	–	–	314

Table 2. Misclassified pixels in noisy image.

Algorithm	Multiresolution	Scalable	Single resolution
Object.	%4.13	%3.04	%8.85
Image	%7.96	%6.20	%18.74

in the range $(-30\ ,\ +30)$ was added to the Table Tennis images, and then different segmentation algorithms were performed. The number of misclassified pixels for the Table Tennis object including arm, racket and ball (11033 pixels) as well as the entire image pixels were counted. The results in Table 2 confirm that the proposed algorithm can deal with noisy images as effectively as multiresolution segmentation and much better than single-level segmentation. This result confirms that the introduced multilevel segmentation algorithm maintains most advantages of multiresolution segmentations over single-level segmentations such as better segmentation of noisy images.

One of the applications of the proposed segmentation algorithm is in object extraction for region-based

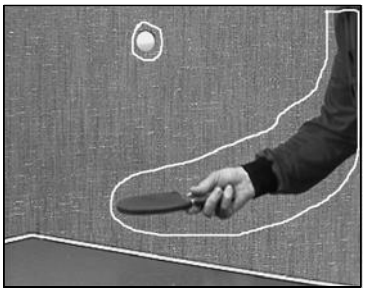


Figure 5. Table Tennis object selection by user;

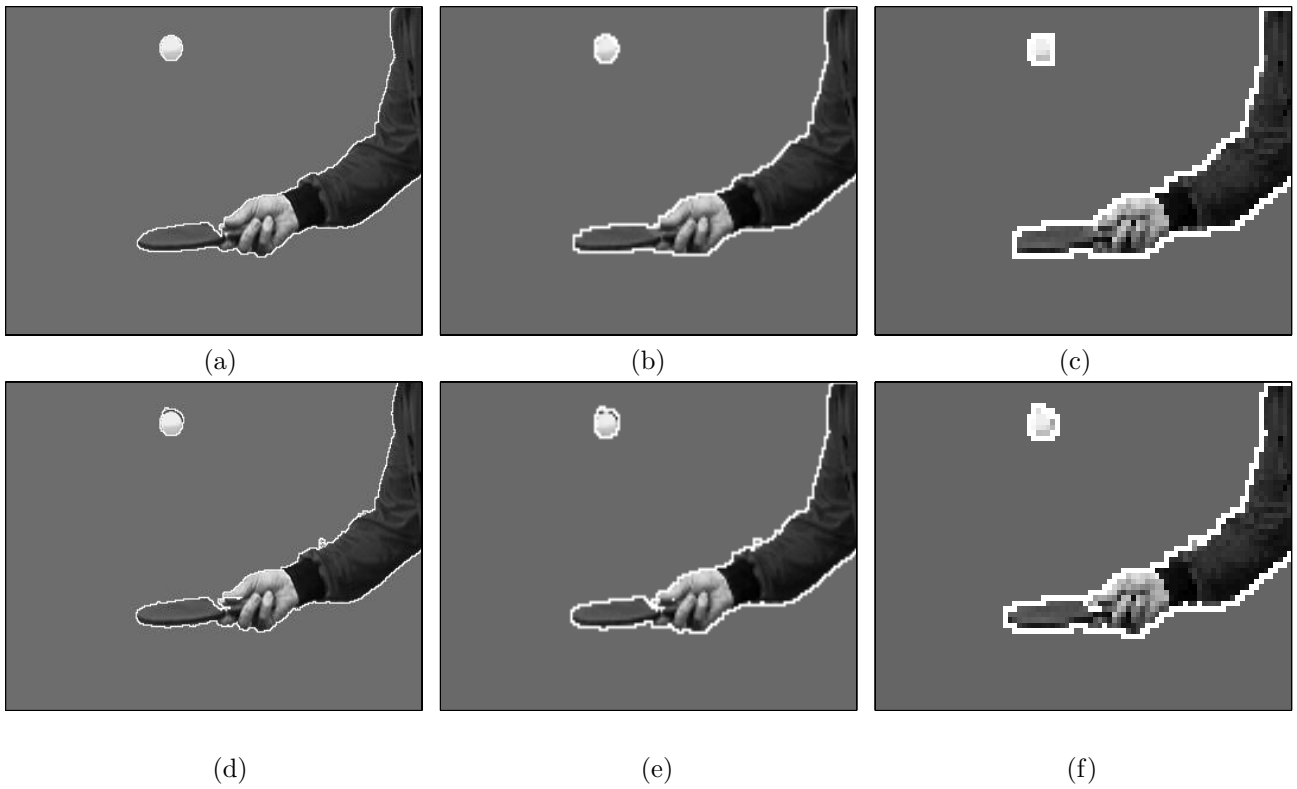


Figure 6. Table Tennis object extraction, objects are extracted form scalable segmentation in the first row and form single-resolution segmentation in the second row; (a) 240×352 ; (b) 120×176 (c) 60×88 (d) 240×352; (e) 120×176 ; (f) 60×88 ;

image/video compression. In many of the semi automatic video object extraction algorithms, in the first frame, the object of interest is determined by user intervention and is tracked in the subsequent frames,[21] .[22] To facilitate the first frame selection, the user can determine the rough boundary of the object of interest through a graphic user interface (GUI) program. Then all the regions with a predetermined percentage of their area inside this closed contour are selected as the region of the extracted object. Joining of all the selected regions creates the final object. As an example, a user has roughly determined the objects of interest in Figure 5. Subsequently the exact borders of the object at different resolutions are determined. The extracted objects by both the scalable segmentation and regular single-resolution segmentation algorithms at three different resolutions are shown in Figure 6. A comparison of the extracted objects confirms the superiority of the scalable segmentation algorithm in a subjective test.

(a) (b) (c) (d)

Figure 7. Frame 34 of qcif size Mother and daughter Image sequence segmentation with $k = 7, 2, 2$ clusters and $\beta = 40$; (a) original grey-level image; (b) regular grey-level single-resolution segmentation; (c) color image of Mother and daughter where U and V are in half resolution; (d) Proposed scalable segmentation.

Table 3. Misclassified pixels in Foreman image segmentation

Resolution	72×88	144×176	288×352
Algorithm A	50	186	786
Algorithm B	123	511	2118
improvement	%59	%64	%63

In a second example, frame 34 of the Mother and Daughter sequence was segmented. The image is in qcif format and is given in the YUV color space. Similar to many other sequences in databases, U and V, the chrominance components, are in half resolution. Regular color-image segmentation needs the information in the same resolution. Therefore, in the first solution, the image was segmented in grey-level space by a single-resolution statistical image segmentation algorithm.[4] The result is shown in Figure 7(b). The left area of the daughter's face has not been well separated from the background because there is not enough grey-level contrast between face and background. The same shortcoming happens for the other grey-level segmentation algorithms except when there is oversegmentation with a large number of detected regions, which is not desired for segmentation applications. To successfully separate object's regions from background, color segmentation was performed as an alternative solution. The proposed scalable segmentation algorithm can perform color segmentation using half-resolution chrominance components. The result of segmentation by the scalable color image segmentation is shown in Figure 7(d). The number of regions in grey-level segmentation is 273 while in color segmentation it is 112, which shows a reasonable color image segmentation algorithm.

In the last example, frame 30 of the qcif sequence foreman is considered. The original image is in YUV format where Y is in full resolution but U and V are in half resolution. In Figure 8(a) the original image can be seen. The image was segmented with the proposed scalable multiresolution segmentation. The result is compared with the other segmentation algorithms. To perform other algorithms, U and V of color components were again projected to full resolution by a $1:4$ pixel transform and the regular single, multiresolution and the proposed scalable segmentation algorithms were performed. The initial segmentation estimation comes from k-means clustering for different channels and the number of classes were chosen as $k = 10, 4, 2$ for the used YUV or RGB color channels, respectively. In the first experiment the components U and V were projected to the next higher resolution and then the proposed scalable and the regular multiresolution segmentation algorithms were performed. The results can be seen in Figures 8(b) and (c). It is clear that in this example regular multiresolution segmentation cannot separate the foreground (foreman) from background regions. This is more pronounced in separating the left area of the hat from the background. Furthermore some other details such as the left eyebrow has not been detected. Similarly, the scalable segmentation using projected U and V to higher resolution could not detect the corner of the hat. The image with real full resolution size of U and V was segmented and will be considered as a ground truth for comparison in the following. The qcif size U and V components were taken from the available YUV, CIF size image sequence. Segmentation by the scalable segmentation which uses real full resolution qcif size U and V

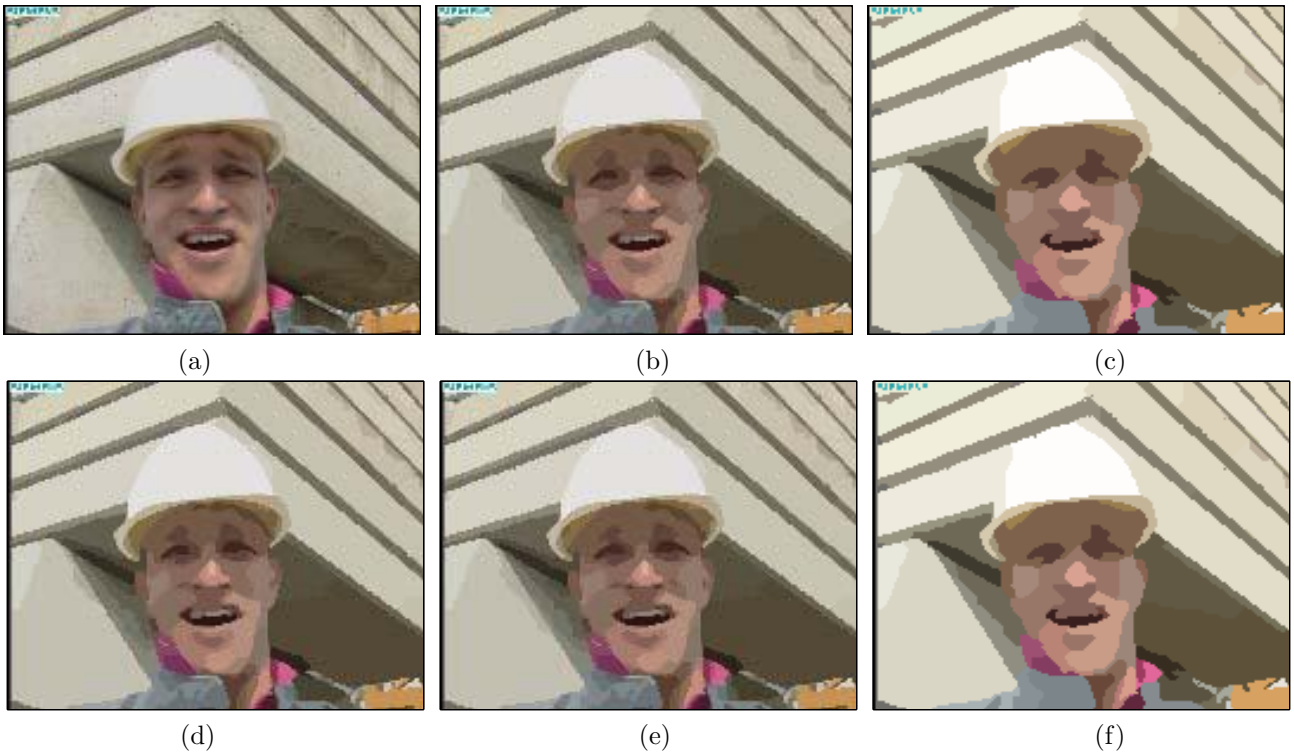


Figure 8. segmentation of frame 32 of qcif size foreman sequence with K=10,4,2 clusters at different channels; (a) original image; (b) scalable segmentation where U and V are projected to higher resolution; (c) Regular multiresolution segmentation (d) scalable segmentation where YUV are in full resolution; (e) scalable segmentation in YUV space where U and V are in half resolution;(f) scalable segmentation in RGB space where components are in full resolution.

are in Figure 8(c). Figure 8(d) shows the segmentation with the proposed algorithm which uses full resolution Y and half resolution U and V components. As can be seen, the proposed algorithm separates the foreground regions from the background successfully, as the scalable algorithm with the full-resolution information does. As a statistical test, scalable segmentation using half resolution U and V (Algorithm A) and scalable segmentation using projected U and V (Algorithm B) are compared with the ground truth. The number of misclassified pixels in the proposed algorithm using half resolution U (Algorithm A) and V is about 30% of the ones of the algorithm which uses the projected U and V components in high resolution (Algorithm B). The numbers of misclassified pixels at different resolutions are shown in Table 3. In the Figure 8(e) the segmented image in the RGB space using the full resolution information is shown. The right and top area of the hat are not separated well. To remedy the problem, we have to increase the number of classes from 10, 4, 2, to 10, 10, 10 classes to separate the hat resulting in the increase of the number of regions which is oversegmentation. Increasing the number of regions will increase the computational complexity of the segmentation algorithm. The number of regions with $K = 10, 4, 2$ is 279 for the proposed algorithm in YUV space and 337 for RGB space which increases to 739 regions for $K = 10, 10, 10$ in RGB space.

5. CONCLUSIONS

We have presented a multiresolution scalable grey-level/color image segmentation algorithm which extracts objects/regions with similar segmentation pattern at different resolutions, which is useful for scalable object-based wavelet coding applications. In addition to scalability, the proposed multi scale analysis incorporate in the objective function of Bayesian segmentation, improves the sensitivity to grey-level variations while maintaining high performance in noisy environments. The extracted objects/regions have better visual quality at different

resolutions. The novel objective function gives flexibility to the proposed algorithm to segment YUV color images where Y is in full resolution but U and V are in half resolution.

REFERENCES

1. F. Pereira and T. Ebrahimi, *The MPEG-4 Book*, Prentice Hall PTR, 2003.
2. B. S. Manjunath, P. Salembier, and T. Sikora, *Introduction to MPEG-7 Multimedia Content Description Interface*, John Willet & Sons, 2002.
3. H. Danyali and A. Mertins, "Flexible, highly scalable, object-based wavelet image compression algorithm for network applications," *IEE Proceedings: Vision, Image & Signal Processing* **151**, pp. 498–510, 2004.
4. T. N. Pappas, "An adaptive clustering algorithm for image segmentation," **40**, pp. 901–914, Apr. 1992.
5. Y.A.Tolias, N.A.Kanlis, and S.M.Panas, "A hierarchical edge-stressing algorithm for adaptive image segmentation," in *Electronics, Circuits, and Systems, 1996. ICECS '96., Proceedings of the Third IEEE International Conference on*, **1**, pp. 199–202.
6. L.Zheng, J.C.Liu, A.K.Chan, and W.Smith, "Object-based image segmentation using dwt/rdwt multiresolution markov random field," in *Acoustics, Speech, and Signal Processing, 1999. ICASSP '99. Proceedings., 1999 IEEE International Conference on*, **6**, pp. 3485–3488.
7. X.Munoz, J.Marti, X.Cufi, and J.Freixenet, "Unsupervised active regions for multiresolution image segmentation," in *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, **2**, pp. 905–908.
8. C.A.Bouman and M.Shapiro, "A multiscale random field model for bayesian image segmentation," *Image Processing, IEEE Transactions on* **13**, pp. 162–177, 1994.
9. M. Comer and E.J.Delp, "Segmentation of textured images using a multiresolution gaussian autoregressive model," *Image Processing, IEEE Transactions on* **8**(3), pp. 408–420, 1999.
10. Z.Kato, J.Zerubia, and M.Berthod, "Unsupervised parallel image classification using markovian models," *Pattern Recognition* **32**, 1999.
11. M. Chang, M. Sezan, and A. Tekalp, "Adaptive bayesian segmentation of color images," *Journal of Electronic Imaging* **3**(4), pp. 404–414, 1994.
12. J. Luo, R. Gray, and H. Lee, "Towards physics-based segmentation of photographic color images," in *Image Processing, 1997. Proceedings., International Conference on*, **3**, pp. 58–61, 1997.
13. J. Gao, J. Zhang, and M. Fleming, "A novel multiresolution color image segmentation technique and its application to dermatoscopic image segmentation," in *Image Processing, 2000. Proceedings. 2000 International Conference on*, **3**, pp. 408–411, 1997.
14. G.Borgefors, G.Ramella, G. di Baja, and S.Svenson, "On the multiscale representation of 2d and 3d shapes," *Graphical Models and Image Processing* **61**(1), pp. 44–62, 1999.
15. A. Mertins and S. Singh, "Embedded wavelet coding of arbitrary shaped objects," in *Proc. SPIE 4076-VCIP'00*, pp. 357–367, (Pert, Australia), 2000.
16. S. Z. Li, *Markov Random Field Modeling in Image Analysis*, Springer Verlag, Tokyo: Japan, seconed ed., 2001.
17. A. M. Tekalp, *Digital Video Processing*, Prentice Hall, USA, seconed ed., 1995.
18. J. Besag, "On the statistical analysis of lattice systems," *Journal of Royal Statistical Society, Series B* **48**(3), pp. 259–279, 1986.
19. T.Meier, K.N.Ngan, and G.Grebbin, "A robust markovian segmenttaion based on highest confidence first (hcf)," in *IEEE Int. Conf. on Image processing*, **1**, pp. 216–219, 1997.
20. C. Rosenberger and K. Chehdi, "Unsupervised clustering method with optimal estimation of the number of clusters: application to image segmentation," in *Proceedings of 15th International Conference on Pattern Recognition.*, **1**, pp. 656–659, 2000.
21. S. Vigus and D. Bull, "Video object tracking using region split and merge and a kalman filter tracking algorithm," in *Processing of International Conference on Image Processing*, **1**, pp. 650–653 vol.1.
22. D. Park, H. Yoon, and C. Won, "Fast object tracking in digital video," *Consumer Electronics, IEEE Transactions on* **46**(3), pp. 785–790, 2000.